Cleaning Up Our Financial System: Combating Money Laundering Using Multiparty Computation

Eniko Kollar , Zekeriya Erkin
Cyber Security Group, Department of Intelligent Systems
Faculty of Electrical Engineering, Mathematics & Computer Science
Delft University of Technology
Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands

Abstract

Globally, it is estimated by the UN that 2% - 5% of the annual GDP is lost to money laundering. Current anti-money laundering efforts are hindered by both the lack of trust between financial institutions internationally and the presence of local privacy regulations like GDPR. This makes it unfeasible to share plain transaction data between financial institutions internationally. Secure multiparty computation is a cryptographic technique that enables a set of parties to interact and compute a joint function of their private inputs while revealing nothing but the output. Thus, MPC has the potential to facilitate greater collaboration between financial institutions and governmental organisations internationally and upscale anti-money laundering efforts securely. In this paper we aim to explore how MPC could be used to improve current anti-money laundering detection techniques. This is done by providing an overview of existing work in the field and proposing a new architecture that could be used to flag suspicious transaction. This architecture presents accounts and transactions as a social network and uses betweenness centrality to identify high-risk accounts. We outline how existing protocols can be used to build such a model and what further properties are to be considered to build even more sophisticated protocols.

1 Introduction

Money laundering is the concealment of the origins of illegally obtained money, typically utilizing transfers involving foreign banks or legitimate businesses [3]. It conventionally consists of 3 stages: *placement*, *layering* and *integration*. During placement dirty money is introduced to the financial system. With layering money is transferred through multiple bank accounts in various jurisdictions to disguise its origins. Finally, during integration, it enters the formal economy by being spent on purchases or invested into legitimate assets.

Investigating the layering process is only possible by a wide international collaboration between financial institutions and governmental organisations. This is because individually they lack the complete picture of the transaction history without helping each other. However, several factors prevent collaboration. These include lack of trust between financial institutions, the obligation to comply with local privacy legislation like GDPR, damage to the reputability of the organisation as well as potential lawsuits in case of a data breach.

Anti-money laundering (AML) efforts help not only to eliminate illegal funds from our financial institutions but to expose the underlying crimes which result in the obtainment of these funds too. Such illegal activities often include but are not limited to drug offences, human trafficking, fraud, gambling, corruption and terrorism. Globally, it is estimated by the UN that 2% - 5% of GDP is lost to money laundering [1]. Financial institutions are not doing enough to combat these serious violations effectively. In the Netherlands alone bank giant, ABN AMRO was fined EUR 480 million [5] in 2021 and ING was fined EUR 775 million [4] 2018 for failing to combat money laundering effectively in particular being unable to unusual transaction patterns.

Current methods to identify suspicious accounts include manual auditing, data mining and machine learning. Traditional investigative techniques like manual auditing are time-consuming and not scaleable, but until recently there existed few alternatives because of the complex nature of money laundering. Recently, data mining techniques like clustering, neural networks [6], using genetic algorithms and heuristics [7] have been introduced. Machine learning methods, like statistical sequential features [8] have been proposed, but a model is ever only as good as its input, so this technique has a limited impact in detecting layering. For both data mining and machine learning methods to become more effective the challenge of fragmented data ownership across many banks and jurisdictions have to be overcome.

The biggest challenges in sharing this data are a lack of trust between institutions as well as privacy regulations. Firstly, financial institutions might be discouraged from sharing data by the possibility of losing competitive advantage compared to other banks leading to financial losses. The second and more prominent challenge is the issue of privacy. Transaction data contains sensitive personal information, thus is protected by privacy regulations like GDPR for the European Union that describes how personal data from the EU has to be processed and stored. Non-compliance to these regulations can result in fines reaching tens of millions of euros [9].

This introduces a further issue with sharing data across many jurisdictions: ensuring that the process is compliant with data protection legislation from all countries. Moreover, sharing data introduces an additional risk of losing customer trust and potentially facing costly lawsuits if the security of the shared data is ever compromised.

Multiparty Computation (MPC) has the potential to offer a solution to the above-mentioned challenges. MPC is a cryptography technique that ensures the protection of the involved parties from each other, rather than from an outside adversary. This is achieved by enabling a set of parties to interact and compute a joint function of their private inputs while revealing nothing but the output [2]. Although research in this field has been going on since the 1970s, it only became efficient enough recently to be used for real-world applications [2]. As of now, MPC has been already implemented for multiple use-cases and is used in real life, such as trying to detect tax fraud [10] in Estonia.

Currently, five Dutch banks Rabobank, ING, ABN AMRO, Triodos Bank, de Volksbank already share data for anti-money laundering under a project called TMNL [11] that does not use MPC. However, this is only possible due to the strong trust between these financial institutions and the regulatory alignment, as all of these banks fall under the same jurisdiction. With the use of multiparty computation, this effort could be further developed into a global collaboration between banks to combat money laundering. MPC, as a privacy-preserving technology has been described as an important tool to enable secure data processing in the *Schrems II* ruling [13] where EU legislators encouraged the use of MPC as a privacy-enhancing technique.

In this paper, we aim to discover what techniques are currently used to combat money laundering. Currently, existing methods that are outlined above are limited by poor data quality due to the lack of collaboration between banks. To enable upscale anti-money laundering efforts, an opportunity for banks to share data and collaborate should be created. We explore how MPC can be used to further these efforts. This is done by outlining an architecture that could be used to identify suspicious bank account holders by enabling collaboration between banks internationally. Hence, we hope to answer the following question as part of our research: *How can multiparty computation be leveraged to combat money laundering by enabling the analysis of data sets from multiple sources?* With this, we hope to contribute by inspiring other researchers, financial institutions and governments to foster closer collaboration to enable anti-money laundering. This is done by proposing to view transactions as a distributed social network and use social network analysis metrics combined with MPC, especially betweenness centrality to identify high-risk accounts.

The remainder of this paper is structured as the following: in the background section anti-money laundering techniques and current MPC use-cases are outlined. In the methodology, the methods of this research are detailed to aid reproducibility. Later in the design section, a new architecture is proposed to use MPC for anti-money laundering. The implications of this architecture are detailed in the discussion session. Afterwards, in the responsible research section, the implications and limitations of this research are outlined. Finally, in conclusion, and future works possible next steps for validation and implementation are discussed.

2 Background

This section aims to summarise currently used anti-money laundering methods as well as a more detailed overview of multiparty computation with an emphasis on methods that have the potential to be used for anti-money laundering.

2.1 Anti-Money Laundering Techniques

AML efforts can take various forms which can be classified into the identification, detection, avoidance and surveillance of money laundering activities [14]. Most current solutions on the market are statistically based on the number of transactions, the amount transferred, mean, standard deviation and thresholds [15]. However, these methods are not suitable to deal with the ever-increasing complexity of ML for example in the case of investment activities that involve completely different characteristics [16]. These statistical methods also produce a large number of false positives and require many man-hours to manually investigate [17]. Data mining (DM) techniques can be effective to reduce the false positivity rate, allow better prioritisation of suspicious transactions and reduce pressure on man-power [17]. Some of the techniques involve: using Bayesian inference to rank ML suspicious transactions, using consolidation and link analysis to determine subgroups and central mules, as well as inter and intro group transaction patterns, deploying regression and case-based reasoning and using support vector machine (SVM), do deal with data of many nationalities [17]. Data mining methods are however limited to the incomplete data, that financial institutions have access to.

2.2 Multiparty Computation

Multiparty computation first proposed by *Yao* [19] is a cryptographic technique. It enables a set of n input parties $p_1, p_2, ...p_n$ with a private set of inputs $x_1, x_2, ...x_n$ to compute the output of a function f on the aggregate of the inputs calculating $f(x_1, x_2, ...x_n)$ without revealing anything but the output [18]. This is done as a way to preserve certain security properties [2]:

- *Privacy:* Besides their own inputs parties should learn nothing but the output of the specified function .
- *Correctness*: No malicious party should be able to influence the output to deviate from the specified function.
- *Independence of Inputs:* Corrupted parties should choose their input independent of honest parties.
- *Guaranteed Output Delivery:* Honest parties should not be prevented to receive the output of the computation by dishonest parties.
- *Fairness:* Corrupted parties should only receive the output of the computation if and only if honest parties receive it too.

The process of computing the output involves three parties: *input parties* that provide the input, *computing parties* that

carry out the computation to calculate the output of function f and *output parties* that receive the output. These roles might be fulfilled by the same set or a different set of parties. For example, input parties might also be the output parties, but it is also possible that input parties are different from the output parties.

Fundamental work in MPC [19–22] since the 80s proved the possibility to securely compute the output of any function in various models. *Yao* developed the *garbled circuits protocol*, which can be utilized in order to compute any discrete function that can be represented as a fixed-size circuit [19]. These general models are however often inefficient and tailored solutions towards specific use-cases can be more optimal. Specific protocols have been developed such as *Private Set Intersection* that allows the efficient computation of the intersection of two sets [23] or *Shamir Secret Sharing* that utilises secret sharing by dividing the secret into a number of shares [24]. Implementations for specific use-cases are also wide-ranging and deployed projects include Danish sugar beets auction, Estonian student graduation rates, wage inequality in Boston and key management all of which are detailed in this [25] very comprehensive paper.

2.3 Using Secure PageRank For AML

In collaboration with ABN AMRO, TNO and Rabobank work on privacy-friendly data analysis for anti-money laundering is already being developed. [26] This collaboration has yielded a model detailed in these [27, 28] papers, which uses multiparty computation to calculate the secure *PageRank* collaboratively. Originally, PageRank has been used by Google to estimate the importance of a page by calculating the number and quality of links that lead to a certain page. Calculating the PageRank value of a bank account can be a good indicator to discover potentially fraudulent accounts. However, it has been proven [27], that PageRank values that are calculated based on the network of multiple banks are more accurate than those calculated merely based on one bank's fragmented data. Thus, the collaboration to calculate these values more collaboratively with the help of additively homomorphic encryption can be valuable [27]. This project is still in progress, although no further work has been published yet. During our interview with one of the authors of this paper, it has been revealed that in the next step of the project another architecture will be researched, where trust flows within the network of bank accounts.

2.4 Rosetta by Roseman Labs

Some MPC engines supporting anti-money laundering have been already implemented. In the Netherlands, Roseman labs have such a solution named *Rosetta* [29]. Rosetta currently applying MPC for pseudonymization and then processing the pseudonymized transaction data. This means that a transaction id from bank A and bank B are pseudonymized with different keys in two different places resulting in two different pseudonyms. Then a central authority can derive a single pseudonym by applying an MPC algorithm to a large number of transactions. For further information, a more detailed product sheet describing Rosetta is available by Roseman Labs on request.

2.5 Secure Privacy-Preserving Graph Algorithms

This paper explores the potential of representing bank accounts and transactions on a graph. Thus, it is important to understand what work has been already conducted to enable the efficient use of Multiparty Computation on graph algorithms. Although real-world applications are yet to be explored, this [30] paper proposes and proves the shortest path and the maximum flow problems can be securely solved using MPC. They explore a secure version of *Bellman-Ford* and *Dijkstra* for the shortest path and *Edmonds-Karp* and *Push-Relabel* for the maximum flow algorithm. These algorithms all build upon the secure arithmetic black-box functionality of *Damgård and Nielsen* [31] which involves assuming a secure implementation of addition and multiplication as well as comparison.

Another solution involves a secure MPC protocol for *k-nearest neighbor search* detailed in this [32] paper. This algorithm has linear computation and communication complexity and is based on a two-party computation model. This model can also be adapted to solve other data mining tasks securely, such as *classification* and *outlier detection* [32].

2.6 Secure Distributed Social Networks

More specifically to graphs, account holders and their transactions can be well understood using *social networks*. There has been already a considerable amount of work done [33,41] to create MPC protocols that allow securely constructing social networks from distributed sources. In the first paper, a general MPC protocol for a distributedly held *isomorphic graph network* was proposed with *unweighted edges*. In the second paper, a similar protocol was proposed that is also secure against the *malicious adversarial model* tolerating less than n/3 corrupt parties.

3 Methodology

Firstly, a literature review was conducted by reading research papers on multiparty computation like [2]. After the basics and the scope of this cryptographic technique were understood, a possible application for the industry was explored.

3.1 Literature Review

The purpose of this study was to provide an overview of the current state of the use of MPC for AML in the financial industry as well as to propose a new use case for using MPC for AML. The best fit to gather and validate this information was a combination of literature study and expert interviews.

We first identified combating corruption as a valuable use-case to research as with combating corruption usually there comes lack of trust and lack of collaboration between various entities to combat it. We felt MPC might the right tool to combat these issues. We further narrowed down our use-case by researching financial crime in general and later money laundering more specifically. We identified different techniques, such as data mining, artificial intelligence and social network analysis. Then, we examined if any work using MPC for AML has been already carried out. Since we concluded that such work, although very limited and recent already exist, we examined the work that has been done. Based on this

research it was concluded that social network analysis has not yet been utilised using MPC for AML. Thus we focused on papers that use secure social network analysis using MPC and papers that use social network analysis for a single, non-distributed graph.

3.2 Expert Interviews

In the second stage of the project expert opinion was seeked out. Below, a list conducted interviews can be seen in Table 1. We conducted 6 interviews with experts in the financial sector, 5 of whom were also experts in multiparty computation. Some interviewees worked at big dutch banks, like Rabobank, ABN AMRO and some were authors of published papers on the topics. We used these interviews to better understand how MPC is currently used in the financial industry and what are the biggest challenges that companies face. We furthermore used these interviews to validate our design ideas. We would like to express our deepest gratitude to all of them for contributing their expertise and time to this project.

Affiliation	Expertise	Date
Rabobank	AML	03/05/2021
Rabobank	MPC	07/05/2021
ABN AMRO	MPC, AML	21/05/2021
Rabobank	MPC, AML	25/05/2021
Roseman Labs	MPC, AML	10/06/2021
Roseman Labs	MPC, AML	16/06/2021

Table 1: List of expert interviews, their affiliations and area of expertise.

4 Design

This section discusses the proposed architecture that facilitates the collaboration of multiple financial institutions and possibly other governmental parties to collaborate to upscale AML efforts.

4.1 Requirements

The proposed architecture has to adhere to the following requirements:

- *Security:* It has to be ensured that no additional data can be recovered from the computation and that the output can not be used to derive additional information to prevent the leakage of sensitive personal data.
- *Performance:* As derived from an interview, for the architecture to be efficient, updating all values should be possible within a matter of days or a few weeks maximum. If this requirement is not fulfilled the architecture might not be of any practical use to further AML efforts.
- *Scalability:* It should be possible to accommodate both networks with many nodes and vertices and preferably both the collaboration of many financial institutions in order to calculate more accurate results.

4.2 Stakeholder Analysis

Main stakeholders in this framework include financial institutions, government institutions and individuals and organisations who hold a bank account. These stakeholders all have different priorities and interests.

Financial institutions want to accurately identify cases of money laundering both to maintain customer trust and to comply with regulations and avoid fines. They also want to maintain client trust and avoid potential lawsuits so the privacy of their client's data is of uttermost importance. Figure 1 shows how banks might share data securely by only exchanging secret shares of the data. Governmental institutions want to ensure that money laundering is either prevented or quickly and effectively detected by financial institutions. They also want to ensure that the privacy regulations are upheld. Finally, bank account holders want to ensure that their sensitive personal data is not compromised.

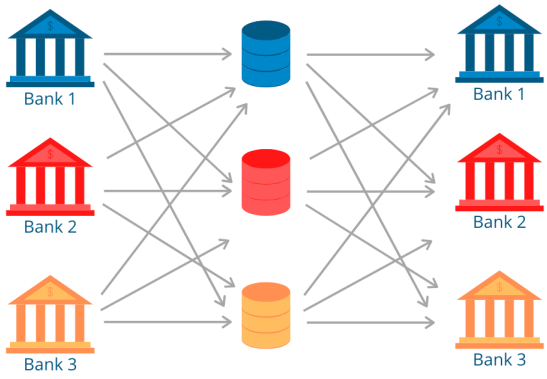

Figure 1: Overview of parties that share data, including three banks as input, computing and output parties.

4.3 Overview of Proposed Solution

In this paper, a *secure distributed social network analysis for money laundering* is proposed. Previous work in using social network analysis [34] proved that representing accounts and transactions as a social network enables the application of several metrics that are significantly correlated with high-risk accounts. One such metric is *degree centrality* which delegates an importance score to each node depending on the total number of links held by that node. It has been concluded [34], that having more central nodes are associated with risky profiles. *betweenness centrality* is a centrality measure that allocates a score based on the number of times that a node is in the shortest path between two nodes. Let $\sigma_v(s,t) \in 0,1$ be the number of shortest paths from vertex s to vertex t running through vertex v. Betweenness centrality [36] is defined as

$$C_B(v) = \frac{\sum_{\substack{s,t \in V \\ s \neq t, s \neq v, t \neq v}} \sigma_v(s,t)}{(|V|-1)(|V|-2)} \quad (1)$$

A high betweenness centrality score has been also identified as being correlated with risky profiles [34]. Typically

a high betweenness centrality score indicates a gatekeeper functionality between two criminal organisations. [35].

In this paper, we focus on creating a secure distributed social network for transaction data and evaluating the risk factor of accounts in this network using betweenness centrality. Betweenness centrality is a good measure to identify money mules and proxies [34, 36]. A high score indicates that an individual account has high influence and authority in a cluster of the network. An example of betweenness centrality values of a small network can be seen in Figure 3.

This has been previous done in this study [34] that has analysed real-world data of Italian financial institutions over a period of 19 months. However, as far as we are aware analysis for anti-money laundering has never been carried out in a distributed, secure manner using MPC. This paper aims to propose an architecture that enables such a framework thus enabling the collaboration between multiple financial institutions to potentially determine betweenness centrality more accurately.

4.4 Data Pre-Processing

Various financial institutions might store their data in different formats and might record different properties of a transaction. For the purpose of this architecture, we assume, that all financial institutions store the following properties of a transaction: source account, destination account, timestamp, amount. Institutions that do not store this data are not able to participate in this collaboration. Furthermore, it is essential, that all data is standardised so that when the social network is being calculated there are no issues such as using different currencies or timestamp format emerges. To prevent data leakage as well as allow space-efficient storage, the transaction data is stored in an edge list format, meaning each transaction represents an edge. This is compatible with the secure *Brandes algorithm* discussed below.

4.5 A Two-Party Architecture: Collaboration Between Two Banks

Brandes algorithm is the most efficient algorithm for computing betweenness centrality in social network analysis. It has a time complexity $O(|V||E|+|V|^2log|V|)$ and $O(|V|+|E|)$ space complexity, where V and E are the number of vertices and edges in a graph, respectively [38]. Today, many other efficient versions of Brandes [39] and other algorithms [40] have been developed to enable efficient computation in large networks. In this architecture we are using a secure version of Brandes algorithm to compute betweenness centrality in the combined network of multiple banks, hence enabling the better identification of high-risk nodes.

Some of the existing research into using MPC for social network analysis are detailed in these [37, 41–43] papers. In 2018 a secure version of the Brandes algorithm was proposed [37] for unweighted networks using multiparty computation. This protocol works in a semi-honest adversarial model in a two-party setting. Meaning that this protocol is assumed to be run by two non-colluding parties who do not deviate from the protocol. With this protocol, if more than two financial institutions wish to combine their network, all financial institutions need to provide shares of their data to two other representative parties. When the data is received these two parties together compute the output for the rest of the parties. Its performance was evaluated in by implementing it in the *Obliv-C framework* for secure computation. Performance was benchmarked using various *ORAM schemes*. Using the *Circuit ORAM* and the *Square-Root ORAM* schemes, the complexity of the protocol is $O(|V||E|log^3|E|)$ and $O(|V||E|^{1.5}log^{1.5}|E|)$ primitive operations respectively [37].

As previously discussed the input of this algorithm is in an edge list format with each transaction representing an edge. The centrality value of each node is computed in stages. The exact protocol can be found in Figure 2 of this [37] paper. The output of this protocol is the list of betweenness centrality scores associated with each node. In the paper [37], the security and correctness properties of the protocol are also proved. Computing betweenness centrality jointly can lead to a more accurate score. This is demonstrated in Figure 3 where betweenness centrality scores are calculated by each bank independently and Figure 3 where this is done jointly.

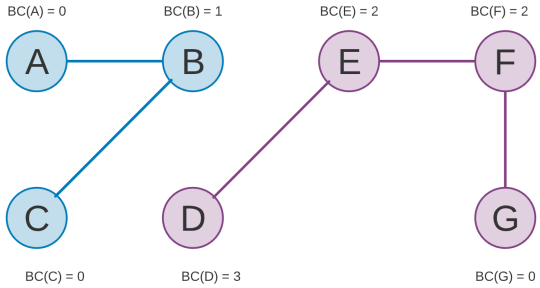

Figure 2: Two banks computing betweenness centrality of their graphs separately.

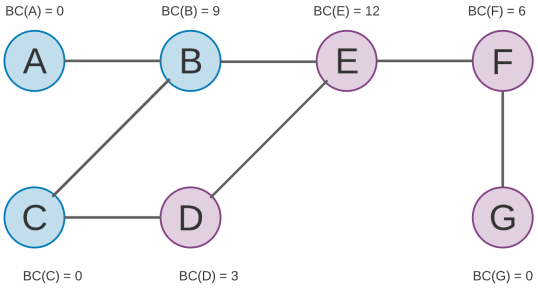

Figure 3: Two banks computing betweenness centrality of their graphs collaboratively using MPC.

4.6 A Multi-Party Architecture: Collaboration Between Many Banks

An alternative protocol was proposed in this [44] paper. Here, both *betweenness centrality* and *closeness centrality* are calculated as the output for $n \geq 2$ parties. Let (v, t) be the

length of the shortest path from vertex v to vertex t in graph G. Closeness centrality [36] is defined as

$$C_C(v) = \frac{\sum_{\substack{t \in V \\ v \neq t}} |V| - 1}{\delta(v,t)} \quad (2)$$

However, using closeness centrality to detect high-risk account involved in money laundering has lead to contradictory results. An analysis of real-world data found that there is no significant correlation between the closeness centrality score and the risk of accounts [34]. However, another study concluded that there exists a correlation and a high closeness centrality score is associated with leadership in a criminal organisation [35]. Thus, more work has to be done to evaluate if closeness centrality is indeed useful in identifying suspicious bank accounts.

Even if closeness centrality does not prove to be a useful measure, a modified version of this protocol can be used to calculate betweenness centrality. This can easily be done by eliminating step 3.2.2.1 in the protocol [34]. Otherwise, the protocol can be used in its original form. First, computing the initial matrix using pseudonyms. Then, computing the set-union of vertices as described in the paper [34] computing the shortest-path matrix using the Floyd-Warshall algorithm. Finally computing closeness centrality and betweenness centrality as described in sections 3.2.2.1 and 3.2.2.2.

For this protocol, the input is the edge list of transaction data and the output is a list of betweenness centrality and closeness centrality score associated with each node.

Like the previous protocol, this protocol is secure for the semi-honest adversarial model as well. However, since it considers a multi-party setting instead of a two-party setting it could have more wide-ranging use-cases as frequently more than two banks want to collaborate.

4.7 Alternative Representations and Optimisations

Another alternative is to utilise *weighted graphs*, instead of unweighted graphs to calculate betweenness centrality. There are many options to decide what the weight of an edge might be: for example, it could be the total number of transactions between the two account or the total amount of money transferred between the two bank accounts. This would significantly reduce the number of edges in the network which could lead to quicker computation using the MPC protocol. However, it introduces an extra step of having to calculate the edge weight for each participating banks on their own before the start of the collaborative calculation.

The above-mentioned protocols also use undirected graphs. It might be interesting to consider using *directed graphs* instead so that the direction of the money flow can be taken into account as well. However, more research needs to be done to see if this would increase the correlation with identifying risky bank accounts.

Moreover, the *temporal significance* of edges could be taken into account as well. Older edges could be either given less weight or completely ignored. This could ensure that the network is not cluttered with old, irrelevant information.

Instead of betweenness centrality as a measure, other measures might be used as well to extract information from the combined network. As we previously measured degree centrality might also be an interesting measure to calculate and has been already proven to be correlated with high-risk accounts. Other measures that might be interesting to explore include:

- *Eigenvector centrality* that also measures a node's influence, but also takes into account how well-connected the neighbours of the selected node are.
- *clustering coefficient* which measures the extent to which nodes in the network are clustered together.
- *structural cohesion* that measures the number of members, that would disconnect a group if removed.
- *propinquity* which measures how closely connected nodes are with other nodes that are geographically close to them.

Lastly, a completely different approach could be taken as well. The combined network could be used to propagate trust through it. For example, each account holder could have a trust score. If a fraudster is identified, mistrust could flow through the network to its neighbours and the neighbours' neighbours. Based on our interviews, this option is already being investigated by the same team that proposed the secure PageRank algorithm, although nothing has yet been published, thus we have chosen not to focus on this architecture.

5 Discussion

In this section, the viability and the wider implications of the proposed architecture are evaluated. These include discussing how the proposed architecture might complement existing AML solutions as well as possible collaboration with companies that have already acquired a related knowledge base. Furthermore, an analysis of its viability and security is provided and an outline of how it might be tested.

5.1 Evaluation of the Proposed Architecture

The first architecture proposed is highly limited when it comes to scalability as it is a two-party computation model. However, it still might be used to enable collaboration between two financial institutions. This can be useful if the banks are strongly connected and a lot of transactions are carried out between the two. It also might be used for scenarios, when a specific account is being investigated with a specific connection, or if a quick analysis is needed and there is no time to compute values across all the participating banks. The second approach is more promising as it allows the collaboration of many different parties. Evaluating the usefulness of jointly calculating betweenness centrality and closeness centrality with real-world data is however needed before any conclusions can be made. Although theoretically, scores should become more accurate under a joint computation, this has to be also proven in practice to justify the risk taken to share sensitive data and the computational power used. Further considerations should be made about the ideal size of the number of collaborating institutions. It might be the case, that computing a graph too big is not efficient compared to the improvement in accuracy in the score.

5.2 Complementing Existing AML Methods

During the interviews, it has been clarified, that flagging a bank account as suspicious from an MPC protocol alone is not sufficient to justify a full investigation on the account. This is because MPC does not provide any other data other than the output of the computation. To justify a full investigation more background data would be needed. This is why we propose to use this MPC architecture as a complement to existing AML techniques rather than its own. Complementing existing techniques with this MPC solution would allow a better prioritisation of which accounts need to be investigated.

5.3 Collaboration with Industry

To test the proposed architecture, real-world data is needed which can be best acquired by collaborating with the industry. Some companies already involved in AML efforts have the potential to become a valuable source of collaboration to implement the proposed architecture or to develop more efficient ones. Knights Analytics [45] is a state of the art company for graph analysis for AML in the Netherlands. They specialise in AML detection for graphs within a single bank. There is a strong case for using MPC to enable combining the graph analysis results of multiple banks.

Another area for potential collaboration is with Moneyou a subsidiary of ABN AMRO. They have between 10 and 100 simple rules that are applied to the transaction flow to filter out suspicious transactions. An example for such a rule would be: is the cash transaction larger than 1000 euros? A decision tree using MPC could be built to data enable sharing across banks. This would be a quick win according to one of the experts interviewed. It might be however computationally expensive to run a decision tree, so optimisations have to be considered.

5.4 Barriers to Adaption

Legislative barriers and lack of understanding of how MPC protocols work have been mentioned most frequently as the biggest barriers to adaption by the experts interviewed. These two issues are strongly intertwined. MPC is a difficult concept to grasp, thus it is hard to persuade management that this would be a valuable technique to devote resources to. Because of the high barrier to entry to understand this technology it is also challenging to gather political support to change any legislation that might prevent collaboration between financial institutions using MPC. Moreover, it can violate consumer trust if data is shared using MPC as the lack of understanding might lead to fearing potential data leaks.

5.5 Viability and Testing

Although a theoretical time and space complexity of the algorithm is provided to establish viability full implementation and performance benchmarking with real-world data is needed. This would allow to better evaluate the computational complexity and accuracy of the proposed architecture.

Experimentation could be done using any open-source MPC engines, for example, *MPyC* [46] the MPC engine developed by the *Eindhoven University of Technology*. This engine also has support for ORM, so it should be possible to create a prototype using it by utilising the constant time algorithm that is provided in Figure 2 of this [37] paper. Although there is an ORM module provided, MPyC is a general engine, so for optimal performance, an engine that gas a specialist package optimised for ORM might be chosen instead. Other existing MPC engines that might be considered include using *Sharemind* [48], *VIFF* [49], *FairplayMP* [50] or *Sepia* [51]. During the interview, it was suggested by experts, that the protocol discussed in the design section looks viable in terms of complexity. They suggested that benchmarking might be done with 10 million transactions for a single bank or to 100 million transactions daily to a country of the size of the Netherlands to establish the performance of the protocol.

5.6 Security

The architecture proposed works under a semi-honest adversarial model, meaning that even corrupted parties correctly follow the protocol. This provides a rather weak form of security and a protocol protecting from malicious adversaries would provide better security by also protecting against parties who do not follow the protocol and who collude with each other. However, a semi-honest adversarial might still be sufficient if we assume that banks that share data this way have no reason to collude with each other, which would mean that data leakages should be avoided.

As security under a semi-honest adversarial model can not be guaranteed, the data might not be considered anonymous thus privacy regulations like GDPR might apply. However, according to an expert, this should not present a problem, because applying GDPR, in this case, should be relatively simple. Even if a privacy impact assessment has to be carried out, privacy officers should find the protection provided by MPC in combination with other privacy-preserving steps sufficient. These techniques might include purpose binding, data minimisation, privacy by design, the fact that the data stays with the owner and that there is no central database where data is stored.

6 Responsible Research

This section outlines the limitations of this research as well as ethical and ethical considerations that have to be taken into account when evaluating this solution.

6.1 Limitations

This investigation has several limitations which have to be communicated clearly in order to accurately determine its implications. Only a high-level overview of the architecture is proposed, which is only validated by expert interviews and rely on protocols that are proposed by very recent papers in 2018 and 2020. This leaves serious gaps in being able to establish the feasibility of the architecture. Implementation, experimentation with mock data and with real data as well as performance benchmarking lacks completely and will have to be established in the future. Since benchmarking for performance and accuracy lacks, it is impossible to conclude if this method is feasible in terms of computational complexity under current technological conditions for practical use. It is also difficult to validate if this method provides too many false positives when flagging suspicious accounts.

Based on our expert interviews, it has been established that under current regulations a solution like MPC would not be sufficient to justify a full investigation on a suspicious bank account. This is because full investigations have to be backed up by data, which MPC protocols can not provide. This means, that this method has the potential to become more of a complement to existing AML methods, rather than a method used on its own.

6.2 Ethical Considerations

There are also several ethical considerations that have to be made when considering this project. Firstly, since the commercial feasibility of MPC is very recent, many of the studies quoted in this paper either only provide a theoretical outline for their solutions or only use dummy data to validate it, but are not yet commercially used solutions. This means, that there is a significant risk, that these methods might not be reliable or feasible in real life, which might compromise the integrity of this study.

Further studies are also needed to establish the security of MPC protocols for graph algorithms. Studies suggest [32], that in some cases it might still be possible to recover unwanted data. There also might be a risk of exposing sensitive information, if a new protocol builds on existing protocols that might be flowed. Thus, it is not clear if the benefits provided by this architecture are worth the privacy risks it introduces. Similarly, since no implementation is provided, more work needs to be done to evaluate if this method is biased towards certain groups for example infrequent account users, and to correct for those potential biases.

It also has to be considered that even if more collaboration thanks to MPC is reached between financial institutions and governmental agencies, some institutions might be still excluded if they don't have the financial resources to provide the necessary computing power that is essential to enable the calculation of trust values. This would risk the potential of furthering existing inequalities, where rich nations might be allowed to luxury to collaborate, while poorer counties are left on their own devices to try to combat money laundering. If the collaboration is effective in tracking suspicious accounts, it might encourage criminal organisations to take their business to poorer counties that do not have the means to collaboratively combat financial crime, thus further increasing existing crime rates.

Lastly, interviews have been conducted only with experts based in the Netherlands which introduces location bias. Financial institutions elsewhere might store data differently, use different analytical techniques or have small scale collaborations with other banks. This study is also limited to considering legal challenges within the EU's GDPR framework thus might fail to consider more stringent local privacy legislation elsewhere.

6.3 Legal Considerations

There are also several legal considerations that need to be taken into account when determining the feasibility of using MPC for AML. Because MPC is a difficult concept to understand, it might become difficult to rally governmental and legislative support behind it, thus lifting the barriers that currently limit its viability. Although MPC is already used by governments for various purposes, for example, *Sharemind* by the Estonian government [48], there is little precedent for using it across many different jurisdictions. Although MPC might be technically compliant with privacy regulations like GDPR there are many unknowns and grey areas when it comes to legal implications.

7 Conclusions and Future Work

In this paper, a privacy-preserving version of calculating the betweenness centrality of a transaction network was discussed. Calculating betweenness centrality is an interesting metric for furthering anti-money laundering as it has been proven to be positively correlated with high-risk accounts [34]. The utilised protocol requires an unweighted transaction graph with bank accounts as nodes and transactions as edges and works under the semi-honest adversarial model. In the future, it would be beneficial to develop a protocol under the malicious adversarial model to provide stronger security guarantees.

In the design section, some alternative ideas are also detailed besides calculating the betweenness centrality of two unweighted graphs. Firstly, the computation of more than two graphs is discussed. Then alternatives like using a weighted graph, a directed graph and considering the temporal significance of the edges are proposed. Alternative social network analysis metrics other than betweenness centrality are also considered. Lastly, a completely different trust-flow based model is mentioned.

With this work, we hope to have inspired wider collaboration between financial institutions across the globe to upscale anti-money laundering efforts by combining their fragmented data. To our knowledge secure social network analysis with MPC has never has been used for this purpose and this is the first time that it is being outlined. The potential for using social network analysis to better understand customer behaviour is vast and is in no way limited to betweenness centrality. We believe that representing data this way would be already beneficial within a single bank itself to better track transactions. However, with the secure collaboration that MPC allows these efforts could make an even bigger difference by enabling international collaboration.

AML efforts have been hindered since criminals started to use layering techniques to wire transfers across many different banks and jurisdictions. This model could facilitate a large scale collaboration to combat financial crime by combining fragmented transaction data on a scale that has never been done before. If money laundering and other financial crime detection improve, it will become ever harder for criminals to legitimise their earnings and to fuel terrorist and human trafficking activities. For the potential to ultimately put a stop to such crimes, we must do everything we can.

References

[1] Overview. (2021). Retrieved from https://www.unodc.org/unodc/en/money-laundering/overview.html

[2] Lindell, Y. (2020). Secure Multiparty Computation (MPC). Cryptology ePrint Archive. https://eprint.iacr.org/2020/300.pdf

[3] Oxford Languages. (n.d.). Oxford Languages. Retrieved June 12, 2021, from https://languages.oup.com/google-dictionary-en/

[4] Arnold, M. (2018, September 04). ING to pay EUR 775m in money laundering case. Retrieved from https://www.ft.com/content/f3e64e3e-b02b-11e8-99ca-68cf89602132

[5] Reuters, A. D. (2021, April 19). ABN Amro to settle money laundering probe for USD 574 mln. Reuters. https://www.reuters.com/business/abn-amro-settle-money-laundering-probe-574-million-2021-04-19/

[6] Le Khac, N. A., Markos, S., Kechadi, M. T. (2010, April). A data mining-based solution for detecting suspicious money laundering cases in an investment bank. In 2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications (pp. 235-240). IEEE.

[7] Le Khac, N. A., Kechadi, M. T. (2010, December). Application of data mining for anti-money laundering detection: A case study. In 2010 IEEE International Conference on Data Mining Workshops (pp. 577-584). IEEE.

[8] Jing, C., Wang, C., Yan, C. (2019, February). Thinking Like a Fraudster: Detecting Fraudulent Transactions via Statistical Sequential Features. In International Conference on Financial Cryptography and Data Security (pp. 588-604). Springer, Cham

[9] GDPR. (n.d.). GDPR.EU. Retrieved June 10, 2021, from https://gdpr.eu/tag/gdpr/

[10] Bogdanov, D., Jõemets, M., Siim, S., Vaht, M. (2015, January). How the estonian tax and customs board evaluated a tax fraud detection system based on secure multi-party computation. In International conference on financial cryptography and data security (pp. 227-234). Springer, Berlin, Heidelberg.

[11] Over ons. (2020, December 2). Transactie Monitoring Nederland. https://tmnl.nl/wie-is-tmnl/

[12] Koch, K., Krenn, S., Pellegrino, D., Ramacher, S. (2021). Privacy-preserving Analytics for Data Markets using MPC. arXiv preprint arXiv:2103.03739.

[13] European Parliament. (2020, July). The CJEU judgment in the Schrems II case. https://www.europarl.europa.eu/RegData/etudes /ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf

[14] R. C. Watkins et al, Exploring Data Mining technologies as Tool to Investigate Money Laundering. Journal of Policing Practice and Research. Vol. 4, No. 2, 2003, pp. 163-178.

[15] Le Khac, N. A., Kechadi, M. T. (2010, December). Application of data mining for anti-money laundering detection: A case study. In 2010 IEEE International Conference on Data Mining Workshops (pp. 577-584). IEEE.

[16] Le Khac, N. A., Markos, S., Kechadi, M. T. (2010, April). A data mining-based solution for detecting suspicious money laundering cases in an investment bank. In 2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications (pp. 235-240). IEEE.

[17] Gao, Z., Ye, M. (2007). A framework for data mining-based anti-money laundering research. Journal of Money Laundering Control.

[18] Hazay, C., Venkitasubramaniam, M. (2016, August). On the power of secure two-party computation. In Annual International Cryptology Conference (pp. 397-429). Springer, Berlin, Heidelberg.

[19] Yao, A. C. C. (1982). Protocols for Secure Computations (Extended Abstract) 23rd FOCS.

[20] Ben-Or, M., Goldwasser, S., Wigderson, A.: Completeness theorems for noncryptographic fault-tolerant distributed computation. In: STOC, pp. 1–10. ACM (1988)

[21] Chaum, D., Cr´epeau, C., Damgard, I.: Multiparty unconditionally secure protocols. In: STOC, pp. 11–19. ACM (1988)

[22] Goldreich, O., Micali, S., Wigderson, A.: How to play any mental game or a completeness theorem for protocols with honest majority. In: STOC, pp. 218–229. ACM (1987)

[23] Hazay, C., Venkitasubramaniam, M. (2017, March). Scalable multi-party private set-intersection. In IACR International Workshop on Public Key Cryptography (pp. 175-203). Springer, Berlin, Heidelberg.

[24] Shamir, A. (1979). How to share a secret. Communications of the ACM, 22(11), 612-613.

[25] Evans, D., Kolesnikov, V., Rosulek, M. (2017). A pragmatic introduction to secure multi-party computation. Foundations and Trends® in Privacy and Security, 2(2-3).

[26] TNO, Rabobank and ABN AMRO work on privacy-friendly data analysis. (2021, March 31). ABN AMRO https://www.abnamro.com/clearing/en/news/tno-rabobank-and-abn-amro-are-working-on-privacy-friendly-data-analysis

[27] Wiggerman, M. How Google's PageRank inspired us to improve collaboration in fraud detection. Innovating in Cyber Security

[28] Sangers, A., van Heesch, M., Attema, T., Veugen, T., Wiggerman, M., Veldsink, J., ... Worm, D. (2019, February). Secure multiparty PageRank algorithm for collaborative fraud detection. In International Conference on Financial Cryptography and Data Security (pp. 605-623). Springer, Cham.

[29] Roseman Labs - Privacy by design. (2021). Roseman Labs. https://rosemanlabs.com/index.htmlproducts

[30] Aly, A., Cuvelier, E., Mawet, S., Pereira, O., Van Vyve, M. (2013, April). Securely solving simple combinatorial graph problems. In International Conference on Financial

Cryptography and Data Security (pp. 239-257). Springer, Berlin, Heidelberg.

[31] Damgård, I., Nielsen, J. B. (2003, August). Universally composable efficient multiparty computation from threshold homomorphic encryption. In Annual International Cryptology Conference (pp. 247-264). Springer, Berlin, Heidelberg.

[32] Qi, Y., Atallah, M. J. (2008, June). Efficient privacy-preserving k-nearest neighbor search. In 2008 The 28th International Conference on Distributed Computing Systems (pp. 311-319). IEEE.

[33] Kukkala, V. B., Saini, J. S., Iyengar, S. R. S. (2017, January). Secure Multiparty Construction of a Distributed Social Network. In Proceedings of the 18th International Conference on Distributed Computing and Networking (pp. 1-4).

[34] Colladon, A. F., Remondi, E. (2017). Using social network analysis to prevent money laundering. Expert Systems with Applications, 67, 49-58.

[35] Xu, J., Chen, H. (2005). Criminal network analysis and visualization. Communications of the ACM, 48(6), 100-107.

[36] Kerschbaum, F., Schaad, A. (2008, October). Privacy-preserving social network analysis for criminal investigations. In Proceedings of the 7th ACM workshop on Privacy in the electronic society (pp. 9-14).

[37] Kukkala, Varsha Bhat, and S. R. S. Iyengar. "Computing betweenness centrality: An efficient privacy-preserving approach." International Conference on Cryptology and Network Security. Springer, Cham, 2018.

[38] Bhardwaj, S., Niyogi, R., Milani, A. (2011, June). Performance analysis of an algorithm for computation of betweenness centrality. In International Conference on Computational Science and Its Applications (pp. 537-546). Springer, Berlin, Heidelberg.

[39] Brandes, U. (2001). A faster algorithm for betweenness centrality. Journal of mathematical sociology, 25(2), 163-177.

[40] Lee, M. J., Lee, J., Park, J. Y., Choi, R. H., Chung, C. W. (2012, April). Qube: a quick algorithm for updating betweenness centrality. In Proceedings of the 21st international conference on World Wide Web (pp. 351-360).

[41] Kukkala, V. B., Saini, J. S., Iyengar, S. (2020). Secure multiparty computation of a social network.

[42] Kukkala, V. B. (2018, April). Privacy Preserving Distributed Analysis of Social Networks. In Companion Proceedings of the The Web Conference 2018 (pp. 873-877).

[43] Dong, W., Dave, V., Qiu, L., Zhang, Y. (2011, April). Secure friend discovery in mobile social networks. In 2011 proceedings ieee infocom (pp. 1647-1655). IEEE.

[44] Kerschbaum, F., Schaad, A. (2008, October). Privacy-preserving social network analysis for criminal investigations. In Proceedings of the 7th ACM workshop on Privacy in the electronic society (pp. 9-14).

[45] Knights Analytics. (n.d.). Knights Analytics. Retrieved June 15, 2021, from https://www.knightsanalytics.com/

[46] Eindhoven University of Technology. (n.d.). MPyC. MPyC by Einhoven University of Technology. Retrieved June 16, 2021, from https://www.win.tue.nl/ berry/mpyc/

[47] Molloy, I., Chari, S., Finkler, U., Wiggerman, M., Jonker, C., Habeck, T., ... van Schaik, R. (2016, February). Graph analytics for real-time scoring of cross-channel transactional fraud. In International Conference on Financial Cryptography and Data Security (pp. 22-40). Springer, Berlin, Heidelberg.

[48] Bogdanov, D., Laur, S., Willemson, J. (2008, October). Sharemind: A framework for fast privacy-preserving computations. In European Symposium on Research in Computer Security (pp. 192-206). Springer, Berlin, Heidelberg.

[49] Geisler, M. (2010). Cryptographic protocols: theory and implementation. PhD Thesis, University of Aarhus.

[50] Ben-David, A., Nisan, N., Pinkas, B. (2008, October). FairplayMP: a system for secure multi-party computation. In Proceedings of the 15th ACM conference on Computer and communications security (pp. 257-266).

[51] Burkhart, M., Strasser, M., Many, D., Dimitropoulos, X. (2010). SEPIA: Privacy-preserving aggregation of multi-domain network events and statistics. Network, 1(101101).